ABSTRACT OF THE DISCLOSURE

A method for providing name-face/voice-role association includes determining whether a closed captioned text accompanies a video sequence, providing one of text recognition and speech to text conversion to the video sequence to generate a role-name versus actor-name list from the video sequence, extracting face boxes from the video sequence and generating face models, searching a predetermined portion of text for an entry on the role-name versus actor-name list, searching video frames for face models/voice models that correspond to the text searched by using a time code so that the video frames correspond to portions of the text where role-names are detected, assigning an equal level of certainty for each of the face models found, using lip reading to eliminate face models found that pronounce a role-name corresponding to said entry on the role-name versus actor-name list, scanning a remaining portion of text provided and updating a level of certainty for said each of the face models previously found. Once a particular face model/voice model and role-name association has reached a threshold the role-name, actor name, and particular face model/voice model is stored in a database and can be displayed by a user when the threshold for the particular face model has been reached. Thus the user can query information by entry of role-name, actor name, face model, or even words spoken by the role-name as a basis for the association. A system provides hardware and software to perform these functions.